META-ANALYSIS OF RANDOMIZED RESPONSE RESEARCH: 35 YEARS OF VALIDATION STUDIES

Gerty Lensvelt-Mulders[1]
Utrecht University, the Netherlands
Joop Hox
Utrecht University, the Netherlands

The Randomized Response Technique (RRT) is a survey method especially developed to ensure the privacy of respondents when studying sensitive issues. In this paper we present the results of a meta-analysis on randomized response validation studies (1965-2000). The focus of this meta-analysis was on the performance of RRT's compared to other, more conventional, data collection methods, and on the question which RRT produced the most valid results. We used a multilevel approach, with a two-level hierarchical model, to do the meta-analysis. RRT produced more valid population estimates for sensitive topics compared to other data collection methods, and the 'Two unrelated questions'-technique proved to be the most valid method. One caution has to be made, only seven strong validation studies were found. In the near future the analysis will be repeated, including all randomized response studies that compare different data collection methods, but lack an external criterion.

Key words: Randomized response technique, sensitive questions, multilevel analysis, Warner.

1 INTRODUCTION

[1] E-mail: G.Lensvelt@fss.uu.nl and J.Hox@fss.uu.nl

Large groups of respondents experience research on sensitive topics as threatening. This can be because it is intrusive to their private sphere, there can be the threat of sanctions for deviant behaviour, or the threat of scrutiny, for instance in organisational research (Lee, 1993). Respondents are known to give more incorrect answers when sensitive questions are asked, than when they are confronted with innocuous questions (Lee, 1993; Rasinski, et al., 1999). Also, due to the enormous increase in survey research people are confronted with (marketing agencies, government, universities), people generally become more worried about the protection of their privacy. This has its effects on response. The non-co-operative group seems to be growing, resulting in increasing problems with non-response and answer distortion (Clark, and Deshairnais, 1998; Groves, 1989; Hox, and De Leeuw, 1994).

Randomized response technique

To tackle the problem of anticipated privacy violation, Warner (1965) introduced a new research technique, the Randomized Response Technique (RRT). He developed a survey method that guaranteed privacy and that had therefore the potential to overcome the reluctance of subjects to reveal sensitive or probably harmful information about themselves (Chaudurhi, and Mukerjee, 1988; Fox, and Tracy, 1986). The crux of the randomized response method is that people answer one out of two questions, selected by a randomizing device. So the interviewer can never know what question is answered by the respondent, guaranteeing respondents privacy.

Warner (1965) proposed a randomization technique now known as the 'Related question'-method. Respondents have to choose with the aid of a randomizing device, one of these two statements:

- I have sensitive attribute A (selected with probability p)
- I do not have sensitive attribute A (selected with probability 1-p)

Without revealing to the interviewer which statement was selected the respondent answers according to his status on attribute A. Elementary probability theory can be used to get a bias free estimate ($\hat{\pi}$) of the proportion of attribute A in the population.

$$\lambda = p\lambda + (1-p)(1-\lambda)$$

$$\hat{\pi} = (\hat{\lambda} + p - 1) / (2p - 1)(p \neq .5)$$

With a sample variance of:

$$\text{var}(\hat{\pi}) = [\pi(1-\pi)/n] + [p(1-p)/n(2p-1)^2]$$

The first paper on randomized response technique was followed by a spree of research, especially on statistical improvements of the method. This was necessary because Warner's method suffers from large standard error, causing low power. To be as powerful a technique as direct questioning, Warner's methods needed as much as 4 times the number of respondents. The best known improvements are the Unrelated questions-method (Greenberg, et al. 1969; Horvitz, Shah, and Simmons, 1967), and the 'Forced response' or 'Contamination'-design by Boruch (1971). Many other statistical improvements of the randomized response technique followed, resulting in methods to solve quantitative problems, lowering standard errors, increasing power, and to make the method applicable for different research settings (i.e. telephone and mail surveys). For a detailed overview on randomized response designs the reader is referred to Fox and Tracy (1986) or Chaudurhi and Mukerjee (1988).

Progress 1965-2000

Since 1965 six review articles have been published. Two kinds of review articles can be distinguished, statistical reviews, with the emphasis on statistical procedures and developments (Antonak, and Livneh, 1994; Chaudhuri, and Mukerjee, R., 1988; Fox, and Tracy, 1986) and comparative reviews, with an emphasis on comparing different survey techniques/methods and their results (Nederhof, 1985; Scheers, 1992; Umesh, and Peterson, 1991).

The main findings from these reviews can be summarised as following.

- Very few strong validation studies, with external criteria to test population estimates, exist. Umesh and Peterson (1991) found seven studies, we found

three more in nine extra years of research[2]. When only those studies are taken into account that compare the results of randomized response techniques with external criteria, seven studies remain (Horvitz, Shah, and Simmons, 1967; Kulka, Weeks, and Folsom, 1981; Lamb, and Stem, 1978; Locander, Sudman, and Bradbury, 1976; Tracy, and Fox, 1981; van der Heijden, et al., 1998, 2000).

- Results are very contradictory. Especially when the 'more is better adagio' is used as a an indication of superiority of the randomized response technique over other methods of data sampling, the superiority of the randomized response technique is as often confirmed as it is not. For instance, on the question 'Have you ever taken something from a shop, that was worth over 50 dollar' Beldt, Daniel, and Garchia (1982) found no indication that the randomized response technique should provide better estimates than direct questioning, where Wimbush and Dalton (1997) found positive results for the same item, their randomized response technique performed better then direct questioning and self administered questionnaires.
- The features of procedures, respondents, sensitive topics, and different randomized response techniques that influence a respondent's attitude towards truthful answering are an underexposed aspect in studies on randomized response (Lensvelt-Mulders and Hox, in preparation). For instance the number of different topics studied with the aid of a randomized response technique is immense. This also counts for the number of different populations, varying from students convenience samples to full population samples. It is difficult to generalise results across so divers topics or populations.

The overall conclusion should be that 35 years of research have not lead to consensus in the field, nor to the description of best practices. Although many statistical elaborations are made to enhance the reliability of the method, and not withstanding, or maybe because of, the fact that many different randomized response procedures and designs are developed, the diversity of outcomes still limits the usefulness in research practice. This is exactly the type of scientific problem that can be clarified with the aid of a meta-analysis. Such meta-analysis is the recommended instrument to bring some order in the contradictory body of findings in randomized

[2] On-line search was done in the following data-bases: Psychological Abstracts, ERIC, Medline, Sage-publications CD-ROM, The SRM-database, and the CIS, pending the research the reference list was supplemented with studies located from reference sections of retrieved studies and a call for unpublished studies was included on the SMSR-net (Survey Methods Research and Statistics)

response studies (Cooper, and Hedges, 1992; Wortman, and Bryant, 1985).

Purpose of the study

This paper presents the results of a multilevel meta-analysis of the strong validation randomized response studies. Two questions will be addressed in the current study:

- Do randomized response techniques produce more reliable population estimates then more conventional data collection methods?
- Which randomized response technique produces the most reliable population estimates?

2 META-ANALYSIS: DESIGN AND METHOD

Meta-analysis is a relatively new but increasingly important research approach to integrate the accumulating knowledge in a research field. It gives the researcher the opportunity to measure the average effect size of a manipulation across multiple studies (Wortman, and Bryant, 1985). It differs from the qualitative review tradition in so far that it is quantitative directed. A meta-analysis report takes the form of an empirical article, with the same sequence of procedures with respect to problem definition, data sampling and statistical analysis (Cooper, and Hedges, 1994; Schwartzer, 1989).

The rationale behind meta-analysis is to provide a measure for the overall effect of a manipulation across studies. To estimate such an overall effect the first question concerns the homogeneity of the results of separated studies. After all, only if all differences between studies can be explained as the mere effect of sampling variation an overall effect can be estimated. When results across studies are not homogeneous the focus of the analysis will become the explanation of the between studies variance with the aid of moderator variables. In this study a multilevel approach to meta-analysis is used to analyse the randomized response studies (Hox, 1995, in preparation, Hox and de Leeuw, 1994). We used a basic two level hierarchical regression model, with data collection method or randomized response techniques nested in studies. Two models were tested. A null model (M0) indicating

an intercept only model, with no explanatory variables, or, in meta-analysis terms, a model that can be used to test the results of different data collection methods are homogeneous across studies. And, when the results are heterogeneous across studies, a conditional model (M1) is tested, with data collection methods or randomized response techniques to explain between study differences.

3 ANALYSIS

The literature search produced 7 studies in which the results of a randomized response study were compared to external criteria: Horvitz, Shah, and Simmons, 1967; Kulka, Weeks, and Folsom, 1981; Lamb, and Stem, 1978; Locander, Sudman, and Bradbury, 1976; Tracy, and Fox, 1981; van der Heijden, et al., 1998, 2000. When two or more articles were based on the same data set, the results were treated as one study. The studies by Van der Heijden and co-workers are therefore collapsed into one study. This leaves us with a data set containing six studies that validated the research results with external criteria.

The studies came from 5 different international journals, and one working paper of the Triangle Research Institute (North Carolina). The sensitive topics under investigation were: university exams, bankruptcy, social security fraud, being arrested, driving under influence (DUI), and having a baby outside marriage. Five studies were conducted in the US and two studies in the Netherlands. The data collection methods used in these studies were Telephone interviewing, Self administered questionnaires (SAQ), Computer assisted self administered interviews (CASI), face to face interviews, and different Randomized response techniques. The randomized response technique used in the studies were the Forced response-method, Two unrelated questions-method, the quantitative method, the Two-stage sampling method, and Kuk's card method.

Data Quality

Studies had to meet minimally three methodological requirements to be included into the meta-analysis. The estimates for 'percentages wrong answers' and their standard

errors (s.e.), or sufficient statistics to compute both, should be given for all groups. Standard errors for the conventional data collection or control groups were calculated as the sqr of pq/N. For the randomized response designs the standard errors were calculated as the sqr of the variance, which was computed according to the demands of the randomized response method that was used. When the equations for the calculation of the population estimate were not given in the article, we followed Fox and Tracy (1986). The article of Horvitz et al. (1967) did not include the information necessary to compute the s.e. of the 'percentages wrong answers' for the randomized response conditions. The s.e. was therefore estimated as .26, the highest s.e. found in the other studies. This is a conservative estimate (mean s.e. = .14), giving the Horvitz study only a small weight in the analysis.

Because we had only seven studies, it was decided to create a new condition, 'conditions within studies' at the highest level for the analysis of the null model.

4 RESULTS

Analysis of different data collection methods

Five methods for sensitive data collection are distinguished, the randomized response technique, face to face interviewing, self administered questionnaires (SAQ), telephone interviewing, and computer assisted self interviewing (CASI).

First we tested the heterogeneity of data collection methods (Table 1, M0) across studies. The results indicated that the data were heterogeneous across studies, as shown in a significant Z-score ($Z = 3.65$, $p = .0001$). Heterogeneous results imply that differences in results across studies can not be explained by sampling variation between studies alone. Secondly, data collection methods were entered into the regression equation (Table 1, M1). As one can see in Table 1, the overall difference between M0 and M1 was small and not significant (Δ-deviance = 2,906). But there were significant differences between data collection methods as indicated by a significant Chi-square ($\chi^2 = 8.649$, $df = 4$, $p = .009$). A second model (M2) compared randomized response techniques with all other data collection methods (intercept added). The difference between randomized response methods and other data collection methods was significant and M2 the better model (Δ-deviance = 32,795).

The randomized response method produced more valid population estimates than the other data collection methods, indicated by the small B-weight, of .011. The other data collection methods produced less valid population estimates, where CASI revealed the highest B-weight.

One restriction has to be made, the sampling variance of the no intercept models as rather high, meaning that there is a large variance between data collection methods and across randomized response methods mutually.

Table 1. Results of the multilevel analysis

all data collection methods				only Randomized response methods		
Step	M0	M1	M2	Step	M0	M3
Intercept	-.230	--	-.368	Intercept	-.036	--
Randomized		.011	-.310	Forced		-.428
Response		(.103)	(.118)	response		(.127)
Telephone		-.319		Quantitative		-.120
		(.233)				(.329)
SAQ		-.344		Two-stage		-.260
		(.234)		sampling		(.362)
CASI		-.810		Two		.071
		(.403)		unrelated		(.126)
Face		-.348		Kuk's card		-.510
interview		(.121)				(.273)
Sampling variance	1	1	1		1	1
Between cond. Var	.121	.158	.104		.255	.068
Deviation	36.129	33.223	29.349		28.539	10.881

Abbreviations: s.e. in parentheses

Analysis of randomized response methods

Five randomized response methods are used in the validation studies: The Forced response-method, Two unrelated questions-method, the Quantitative method, the Two- stage sampling method, and Kuk's card method.

The results of the first analysis (Table 1, M0) indicated that the data were heterogeneous across studies (Z = 1.9429, p = .026). Then the different forms of the randomized response method were entered as dummies into the regression equation (M3). Differences between randomized response methods were almost significant (χ^2 = 9.747, df = 4, p = .053). M3 was a better model then M0, with a significant Δ-deviance of 17,658. The Two unrelated questions-method produced the most valid population estimates (B-weight = .071) across studies. The highest B-weight is for Kuk's card method.

5 DISCUSSION AND CONCLUSIONS

Randomized response techniques produced more reliable population estimates then the other data collection methods, and the two unrelated questions-method (Greenberg et al., 1969) out-performed the other randomized response methods. That are in short the results of the meta-analysis on seven studies that compared the results of different methods of data collections with an external criterion. The issues under investigation were all sensitive by nature, like issues in the area of sexual behaviour, police arrests, social security fraud, and illegal abortion.

Because only seven studies met the demands of (1) having an external criterion, and (2) the comparison of randomized response techniques with other methods of data collection, it was not possible to investigate the influence of other mediating variables. There are signs, that the content of the sensitive issues, features of the respondents like trust in, and understanding of, the randomized response method, the extend to which the sensitive questions lead to social desirable answers, and the features of the randomisation procedure, all contribute to the robustness of the randomized response method. Firstly, the sample variance between randomized response methods is high, thus the variance across different randomized response methods is large. For research practice this means that randomized response methods are not very robust. Randomized response research is sensitive to small variations in the design and procedures of the study. Secondly there are the results of a preliminary

regression analysis, with population estimates as dependent variables and content of the sensitive topic and social desirability distortion of a topic as explaining variables. The differential content of sensitive topics explained 15 % of the total variance in population estimates, when also social desirability was entered into the regression equation 32% of the total variance could be explained. Although these results are highly preliminary, they do show the importance of moderating variables. Further research on the robustness of different randomized response techniques will help us to explain the noise between studies, like the different outcomes of the Beldt, et al. (1982) and the Wimbush, and Dalton (1997) study.

Salient results of this study are there for CASI as a method for data collection, and Kuk's card method as a method for randomized response. CASI, a computer assisted survey method, is usually a very reliable method (Weisband, and Kiesler, 1996). In this meta-analysis only one study used CASI as method of data collection. In this study the CASI produced even lower estimates than the face to face interview condition. The reason for this salient results could have been the difficult group of respondents. People who received unemployment benefit or welfare, and had a lower educational level than that of the general population (Van der Heijden et al., 1998, 2000).

Also Kuk's card method was used in only one study. Within this study it out-performed the forced response method in two ways. It produced more valid population estimates, that were closer to the external criterion (van der Heijden, et al., 1998, 2000), and respondents reported more trust in, and understanding of Kuk's card procedure compared to other methods of data collection (Landsheer, et al., 1999).

It is important to keep in mind that validation studies with external criteria, on sensitive topics are difficult to develop. For instance, there are ethical issues considering the privacy of respondents that have to be solved. Respondent groups have to come from known registers, like police files (DUI, Fraud), hospital registers (abortion), and bankfiles (bankruptcy). The co-operation of these organisations is therefore needed, making this research more complex. We consider these the reasons why only nine validation studies are carried out between 1965 and 2000. It is therefore a lost opportunity when not all proceedings and features of the randomized response study, including statistical considerations, are written down carefully in research records, so that others can repeat the study or use the results in a meta-analysis.

The results of this meta-analysis are very promising. In the near future the analysis will be repeated, including all randomized response studies carried out between 1975 and 2000, that compare different data collection methods, but lack an external criterion. These studies will make it possible to gain more understanding of the features of randomized response research that will add to a better procedure and more robust research strategy.

References

Antonak, R. F., & Livneh, H. (1995). Randomized response technique: A review and proposed extension to disability attitude research. *Genetic,-Social,-and-General-Psychology-Monographs, 121,*1, 97-145.

Beldt, S. F., Daniel, W. W., & Garchia, B. S. (1982). The Takahasi-Sakasegawa randomized response technique: A field test. S*ociological Methods and Research, 11,* 101-111.

Boruch, R. F. (1971b). Assuring confidentiality of responses in social research: A systematic analysis. *American Psychologist, 26,* 413-430.

Chaudhuri, A., & Mukerjee, R. (1988). *Randomized response: theory and techniques.* New York: Marcel Dekker.

Clark, S. J., & Desharnais, R. A. (1998). Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods, 3, 2,* 160-168.

Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis.* New York: Russel Sage Foundation.

Fox, J. A., & Tracy, P. E. (1986). *Randomized response: A method for sensitive surveys.* Beverly Hills: Sage Publications.

Greenberg, B. V., Abul-Ela, A. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association, 66,* 243-250.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Horvitz, D. G., Shah, B. V., & Simmons, W. R. (1967). The unrelated question randomized response model. *Proceedings in the Social Statistics Section, American Statistical Association,* 65-72.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publications.

Hox, J. J. (in preparation*). Multilevel analysis of regression and structural equation models.* New Jersey: Lawrence Erlbaum.

Hox, J. J., & De Leeuw, E. D. (1994). A comparison of non-response in mail, telephone, and face-to-face surveys. *Quality and Quantity, 28,* 329-344.

Kulka, R. A., Weeks, M. F., & Folsom, R. E. (1981). A comparison of the randomized response approach and direct questioning approach op asking sensitive survey questions. *Working paper, 'Research Triangle Institute'.*

Lamb, C. W. & Stem, D. E. (1978). An empirical validation of the randomized response technique. *Journal of Marketing Research, 15,* 616-621.

Landsheer, J. A., van der Heijden, P., & van Gils, G. (1999). Trust and understanding, two psychological aspects of randomized response. *Quality and Quantity, 33,* 1-12.

Lee, R. M., (1993). *Doing research on sensitive topics*. London: Sage Publications

Locander, W, Sudman, S, & Bradsburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of American Statistics, 71,* 354, 269-275.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15,* 3, 263-280.

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. *Applied Cognitive Psychology,* 465-484.

Scheers, N. J. (1992). A review of randomized response techniques. *Measurement-and-Evaluation-in-Counseling-and-Development, 25,* 27-41.

Schwartzer, R. (1989). *Meta-analysis programs.* Berlin.

Tracy, P. E., & Fox, J. A. (1981). The validity of randomized response for sensitive measurements. *American Sociological Review, 46*, 187-200.

Umesh, U.N., & Peterson, R.A. (1991). A critical evaluation of the randomized response method. *Sociological Methods and Research, 20,* 1, 104-138.

Van der Heijden, P. G. M., Van Gils, G., Bouts, J., & Hox, J. (1998). A comparison of randomized response, CASAQ, and direct questioning; eliciting sensitive information in the context of social security fraud. *Kwantitatieve Methoden, 19,* 15-34.

Van der Heijden, P. G. M., Van Gils, G., Bouts, J., & Hox, J. (2000). A comparison of randomized response, CASI, and face to face direct questioning; eliciting sensitive

information in the context of welfare and unemployment benefit. S*ociological Methods and Research,28,* 4, 505-537.
Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association, 60,* 63-69.
Weisband, S, & Kiessler, S. (1996). Self disclosure on computer forms: Meta-analysis and implications. in Chi'96 electronic proceedings.
http://www.al.arizona.edu/~weisband/chi/chi96.html
Wimbush, J. C., and Dalton, D. R. (1997). Base rate for employee theft: Convergence of multiple methods. *Journal of Applied Psychology, 82,* 756-763.
Wortman, P. M., & Bryant, F. B. (1985). School desegregation and black achievement: an integrative review. *Sociological Methods and Research, 13*, 289-324